

Tarea 2: HDFS y MapReduce

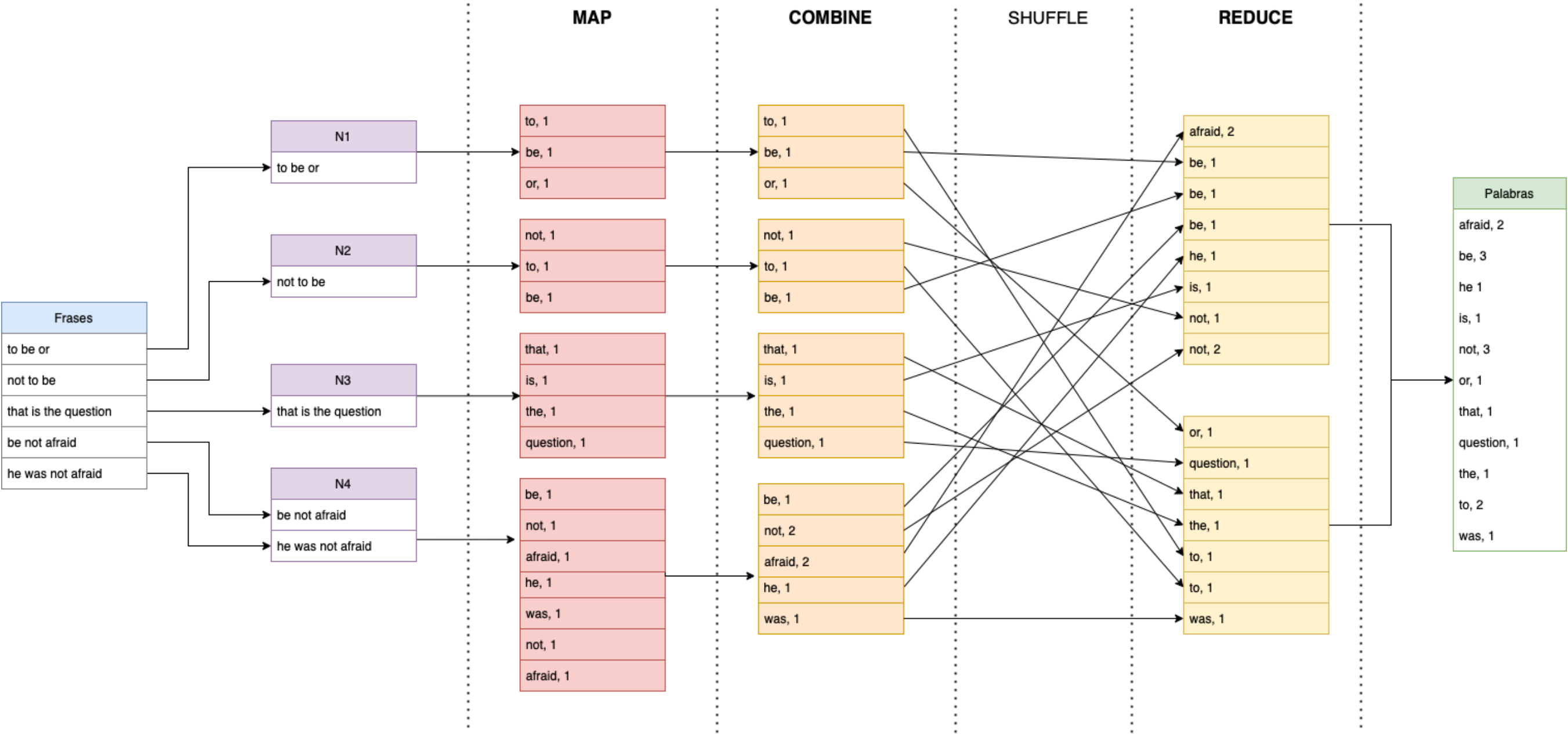
Fragoso Alvardao Daniel | 316049054

1. Dibuja el esquema MapReduce que permita contar la frecuencia de palabras de los siguientes fragmentos:

| F1: to be or | | :--- | | F2: not to be | | F3: that is the question | | F4: be not afraid | | F5: he was not afraid |

Se indicó usar 4 nodos para mapeo y 2 para reducción. En el esquema se debe mostrar claramente cada uno de los pasos y resultados para cada uno de los nodo.

En este caso vamos a iniciar con una lista que contiene todas las frases. El paso inmediatamnete posterior a la entrada es la división de las frases, en este caso se dividiran en 4, cada sección irá a su respecivo nodo, y se realiza el proceso de **MAP**, posteriormente para intentar simplificar las operaciones se realiza el proceso de **COMBINE**, y se agrupan las respectivas llaves valores en dos grupos, uno para cada nodo desigandos a la reducción, finalmente, se realiza el proceso de **REDUCE** y se genera la salida a partir de ambos conjuntos.



1. Diseña una arquitectura basada en clúster que contenga 4 racks en total, 2 racks deben estar conectados a un switch y los otros 2 racks deben estar conectados a otro switch. Ambos switches deben estar conectados a un switch principal. En la arquitectura diseñada etiqueta: la velocidad de comunicación entre nodos y entre racks, así mismo indica el límite inferior y superior de nodos que debe contener cada rack, y menciona cuáles son los elementos que componen a un nodo.

Recordemos que la arquitectura basada de cluster consta de un conjunto de computadoras independientes, interconectadas entre sí, de tal manera que funcionan como un solo recurso computacional.

Así un **switch** es la froma en que los nodos se comunican, cuya velocidad es de 1Gbps entre cada par de nodos. Mientras que el **rack** es una colección física de nodos en nuestro clúster (entre 16 y 64). Y al mismo tiempo cada **nodo** es simplemnete una computadora, es decir que se compone básicamnete de *CPU*, *Memoria* y *Disco*. Finalmente la conexión entre rack's se da por medio de un switch principal cuya velocidad de conexión es mayor que la usal, de 2 hasta 10 Gbps.

